Exhibit 4

Journal of Psychoeducational Assessment

http://jpa.sagepub.com

Extended Time as a Testing Accommodation for Students With Reading Disabilities: Does a Rising Tide Lift All Ships?

Lawrence J. Lewandowski, Benjamin J. Lovett and Cynthia L. Rogers

Journal of Psychoeducational Assessment 2008; 26; 315 originally published online Apr 9, 2008;
DOI: 10.1177/0734282908315757

The online version of this article can be found at:
http://jpa.sagepub.com/cgi/content/abstract/26/4/315

Published by:
SAGE
http://www.sagepublications.com

Journal of Psychoeducational Assessment
Volume 26 Number 4
December 2008 315-324

10.1177/0734282908315757
http://jpa.sagepub.com
hosted at
http://online.sagepub.com

Extended Time as a Testing Accommodation for Students With Reading Disabilities

Does a Rising Tide Lift All Ships?

Lawrence J. Lewandowski
Syracuse University
Benjamin J. Lovett
Elmira College
Cynthia L. Rogers
Syracuse University

Testing accommodations have become a common component of services for students with disabilities at all levels of education. This study examined the effect of a common testing accommodation—extended time—on the reading comprehension test performance of high school students. Sixty-four students, half of whom had learning disabilities (LDs) in the area of reading, were given the Nelson Denny Reading Comprehension subtest under various time conditions. Nondisabled students benefited more from the extended time than students with LDs did. However, extended time did allow students with LDs to attempt as many questions as their nondisabled peers did under standard time conditions. Implications for future research, as well as policy in this area are discussed.

Keywords: *learning disabilities; test accommodations; extended time; reading tests*

The number of students with disabilities being served in American schools and colleges has risen steadily over the past few decades (Thurlow, Thompson, & Lazarus, 2006), and parallel to this trend has been an increase in "high stakes" testing in educational settings (Katsiyannis, Zhang, Ryan, & Jones, 2007). To ensure that the academic skills of this increasingly diverse body of students can be assessed, students with disabilities are often provided testing accommodations. The argument behind these accommodations is that tests should be fully accessible to students with disabilities and that a test may unfairly penalize students with disabilities because of their deficiencies in ancillary skills that are needed to complete the test in the way that it is typically administered. A testing accommodation allows the students to show their actual skill levels, thereby enhancing the validity of their scores. For instance, giving a Braille form of a mathematics test to students with visual impairments keeps these students from being penalized for deficits in a skill (visual acuity)

Authors' Note: Correspondence concerning this article should be addressed to Lawrence J. Lewandowski, Department of Psychology, Syracuse University, 430 Huntington Hall, Syracuse, NY 13244; e-mail: ljlewand@syr.edu.

that is not part of the construct (mathematical knowledge) that the test is designed to measure. Importantly, accommodations are not supposed to alter the construct being measured by the test, so that scores of standard and nonstandard test administrations are comparable (Thurlow et al., 2006).

Extended time to complete tests has become one of the most requested and granted accommodations for both classroom and "high stakes" tests such as the SAT (Bolt & Thurlow, 2004). A common rationale for this accommodation is that students with certain high-incidence disabilities (e.g., learning disabilities [LDs], attention deficit hyperactivity disorder [ADHD]) have deficits in processing and/or reading speed (for a recent defense of this rationale, see Ofiesh, Hughes, & Scott, 2004). In addition, extended time is a common adjunct to more intensive accommodations, such as reading a test to students and writing down responses for students (Thurlow, Elliott, & Ysseldyke, 2003).

Validating Testing Accommodations

Testing accommodations raise concerns about test score interpretation because they involve breaking the standardization in administration that is the hallmark of most tests (Lee, Reynolds, & Wilson, 2003). Therefore, accommodations are generally deemed appropriate only when they minimize the negative effects of the disability while not significantly affecting the performance of nondisabled examinees (Fuchs & Fuchs, 2001). Tindal and Fuchs (1999) referred to this standard as "differential boost," because the scores of examinees with disabilities should get a greater boost than the scores of nondisabled examinees.

In a review of differential boost research, Sireci, Scarpati, and Li (2005) examined 28 studies, concluding that extended time tends to boost scores of all students, although the boosts are somewhat larger for students with disabilities. Interestingly, these conclusions were drawn in spite of the tremendous range of results reported across the reviewed studies; in certain studies, neither group of participants benefited from extended time accommodations, whereas in other studies, both groups benefited substantially. It is difficult to draw integrative conclusions from these studies, because they utilize different methods, use student groups of different ages, measure different academic skills, and have a variety of design limitations. Of importance, in most of the extant studies, almost all students completed the exam when given extended time, meaning that ceiling effects may explain why extended time reduced any differences between students with and without disabilities (Zuriff, 2000).

A recent investigation by Lewandowski, Lovett, Parolin, Gordon, and Codding (2007) deliberately removed ceiling effects by using a highly speeded mathematics test that no student was able to finish, even with extended time. These investigators found that the extended time helped both students with and without ADHD and that the benefit was actually greater for nondisabled students. They noted that when a test is speeded (as many "high stakes" tests are), and ceiling effects are removed, nondisabled students may benefit more from extended time accommodations than do their peers with disabilities, because the nondisabled students are able to accomplish more work during the additional time.

Although Lewandowski and colleagues' (2007) findings are provocative, they may not extend beyond the ADHD population. Students with ADHD are often quite distractible

(Barkley, 1997), and this may keep them from utilizing the extended time productively. The boost favoring nondisabled students, then, may not be found in students diagnosed with LDs, who constitute the largest population of students given extended time. Students with LDs are generally characterized as showing deficits in such cognitive skills as processing speed and reading fluency (e.g., Mather & Wendling, 2005), leading certain scholars (e.g., Ofiesh et al., 2004) to recommend extended time accommodations. Shaywitz (2003) went even further, arguing that individuals with dyslexia simply cannot take tests without extended time; she compares extra time for individuals with dyslexia to insulin for individuals with diabetes.

Claims such as these motivated the present study. We sought to extend the work of Lewandowski et al. (2007) to individuals with LDs, to see whether these individuals would benefit more than their nondisabled peers from getting additional time to take a common test of reading comprehension. To ensure that the students with disabilities would have deficits relevant to the test, students with reading disabilities were the focus of the study. The number of comprehension items answered correctly was the primary dependent measure, with additional analyses based on number of items attempted and errors. To reduce the potential for ceiling effects, we established a "standard" time for the test at 13 min (rather than the typical 20 min suggested in the test manual), and then added 6.5 min or 50% more time to create the "extended time" condition.

Method

Participants

Sixty-four students in grades 10 through 12 (55% female, 45% male) participated in the study. Participants were recruited from a high school located in a suburban school district in New York State. One of the authors was employed by the school district and sent a letter of participation to the families of 70 students with LDs and 70 peers without disabilities. Thirty-two of the students who chose to participate had been given a LD classification by a multidisciplinary school team based on a psychoeducational evaluation, which also met research criteria for LD in the area of reading, that is, each had an aptitude-achievement discrepancy of at least 16 points and demonstrated impairment in learning not attributed to sensory, psychiatric, or environmental factors. The majority of these students had LDs in additional areas (e.g., written expression or mathematics).

The nondisabled group comprised 20 females and 12 males selected from the same classes as the LD group. The LD group consisted of 12 females and 20 males. The preponderance of males in the LD group is consistent with national trends in diagnosis (Fletcher, Lyon, Fuchs, & Barnes, 2007). Ethnicity for the entire sample was as follows: 77.8% White, 7.9% Latin American, 6.3% African American, and 7.9% Other, with comparable ethnicity data in the LD and nondisabled groups, $\chi^2 = 5.38$, $p = .15$, $\varphi = .29$. The age range for all participants was 15 to 18 years, with a mean age of 16.3 years for both the LD and nondisabled groups.

Measures

Reading comprehension test. The comprehension subtest of the Nelson-Denny Reading Test (NDRT; Brown, Fishco, & Hanna, 1993) was administered to both groups of students to assess reading comprehension skills. This subtest consists of seven reading passages with several comprehension questions following each passage. The NDRT manual (Brown et al., 1993) reports an alternate-form reliability coefficient of .81, but because of the test's speeded nature, no internal consistency statistics are reported. We modified the typical time allotments on the NDRT to control for ceiling effects, making it less likely that examinees would finish without extended time accommodations. For the present study, 13 min was the "standard" time allotment and 19.5 min (50% additional time) was the "extended" time allotment. Previous research on lengthening and shortening time allotments has shown negligible effects on psychometric characteristics such as reliability (Willingham et al., 1988).

General cognitive ability measure. The standard form of the Raven Progressive Matrices test (RSPM; Court & Raven, 1995) was used to estimate students' general cognitive ability levels. The RSPM consists of 60 items, each of which displays a matrix that has room for nine elements, but where one of the elements is missing and the examinee chooses one option from six or eight multiple-choice alternatives that would best fit in the empty spot. Reliability coefficients are generally in the range of .90 to .95, and validity coefficients (correlations with other intelligence tests as well as achievement measures) are generally in the range of .50 to .60 (Bracken & Naglieri, 2003). Scores on this test were used to determine whether our groups differed with regard to general cognitive ability, and the RSPM was chosen for its nonverbal nature, as verbally loaded tasks may underestimate the general cognitive ability of individuals with language-based LDs (Kaufman, 1990).

Reading fluency test. The Reading Fluency subtest of the Woodcock-Johnson Tests of Achievement, Third Edition (WJ-III; Woodcock, McGrew, & Mather, 2001) consists of 98 items that require examinees to read brief sentences and decide whether each is true or false by marking "yes" or "no." There is a limit of 3 min for this task. According to the test manual, this subtest has a test–retest reliability coefficient of .88, which is based on a sample of individuals across all age groups. This test was used to describe the presumed group differences in reading speed, which has been shown to predict comprehension scores (Lewandowski, Codding, Kleinmann, & Tucker, 2003).

Procedure

The measures were administered in the students' high school building to groups of students ranging in size from 4 to 20. Students were given a packet that contained all measures in a prepared order and two pens of different colors. After students gave assent, they completed the NDRT, the W-J Reading Fluency test, and finally the RSPM. It took approximately one hour to complete the session. All sessions were administered by a trained doctoral student, and procedural integrity was monitored by a trained undergraduate student. The undergraduate monitor was trained on all procedures by the experimenter and was then present for three practice-testing sessions. The monitor followed a specific observation

Table 1
Group Differences on Performance and Rated Variables

	Nondisabled		Learning Disability			
Measure	*M*	*SD*	*M*	*SD*	*t*(62)	*d*
Raven IQ	98.76	9.62	97.8	11.16	0.37	0.09
Reading fluency	107.1	9.59	88.88	10.41	7.27***	1.82
13 min correct	17.72	4.99	7.0	2.68	10.71***	2.68
19.5 min correct	26.91	4.45	12.38	4.11	13.58***	3.39
13 min attempt	22.88	5.45	12.0	3.41	9.57***	2.39
19.5 min attempt	35.16	3.17	21.25	5.42	12.53***	3.13
13 min % correct	77.48	12.5	59.2	19.2	4.51***	1.13
19.5 min % correct	76.72	11.47	58.37	13.99	5.74***	1.43

***$p < .001$.

protocol to record the extent to which the experimenter followed the experimental script. Sessions had to follow 95% of steps correctly to be considered valid, and each session met that criterion. The monitor also assisted the experimenter in monitoring attention to the tasks and to the changing of colored pencils. Participants were able to follow these simple instructions in a timely manner.

The administration of the NDRT Comprehension test involved two testing conditions: standard and extended time. Students were instructed on the duration for completion of the test and on the need to change pens during the test. They were told to change pens at 13 min. Students were then given an additional 6.5 min, for a total testing time of 19.5 min. Changed answers and completed answers during the normal and extended time period were identified with the different colored pens used for each timing condition.

Results

Group Comparisons

Group comparisons were conducted with *t* tests on several dependent measures, as seen in Table 1. Results revealed no significant difference in RSPM scores (for the nondisabled group, $M = 98.76$; for the LD group, $M = 97.8$, $t(62) = .37$, $p > .50$, $d = .09$), but a significant difference was found in performance on the Reading Fluency subtest, $t(62) = 7.27$, $p < .001$, $d = 1.82$. The nondisabled students ($M = 107.06$) performed better on this measure than the LD students ($M = 88.88$), which is consistent with the LD diagnostic status.

Differential Boost Analyses

Three 2 × 2 mixed-model analysis of variance (ANOVA) procedures were conducted on the following dependent variables: number of NDRT items answered correctly, number of items attempted, and percentage of items correct (see Table 2 for a tabular summary of the results). All variables for both groups met expectations for homogeneity of variance. In

Table 2
Two-Way Analyses of Variance for Items Correct, Items Attempted, and Percent Correct on the Nelson Denny Reading Comprehension Test

Variable and Source	*df*	Mean Square	*F*	η^2
Items correct				
Group	1	5100.5	178.8**	.74
Time	1	1696.53	290.42**	.82
Group × time	1	116.28	19.91**	.24
Within-group error	62	5.84		
Items attempted				
Group	1	4912.88	149.79**	.71
Time	1	3708.76	488.99**	.89
Group × time	1	73.51	9.69*	.135
Within-group error	62	7.58		
Percent correct				
Group	1	10737.13	29.1**	.32
Time	1	20.21	0.355	.006
Group × time	1	0.03	0.001	.00
Within-group error	62	57.01		

*$p < .05$. **$p < .01$.

each ANOVA, group (LD, no disability) was the between-subjects variable and time (standard time, extended time) was the within-subjects variable. If differential boost were present, the extended time accommodation would have a greater positive effect on the performance of students with LDs. We present an effect size measure (η^2) for each ANOVA test statistic, and we recommend Cohen's (1988) interpretation guidelines, in which .01 indicates a small effect, .06 indicates a medium effect, and .15 indicates a large effect. When we make a simple two-group comparison, we report a simpler measure of effect size, the standardized mean difference (d), and we again recommend Cohen's rough interpretation guidelines (.25 = small, .50 = medium, .80 = large).

Items Correct. With regard to the number of NDRT items answered correctly, a significant interaction between group and time was found, $F(1, 62) = 19.91$, $p < .001$, $\eta^2 = .24$. Analysis of the interaction indicated that the nondisabled students benefited more from extended time than did the LD students. The nondisabled students outperformed the LD students at both standard time, $t(62) = 10.71$, $p < .001$, $d = 2.68$, and extended time, $t(62) = 13.58$, $p < .001$, $d = 3.39$. The difference in correctness of items between the two groups became larger as time progressed. Therefore, differential boost favoring students with disabilities was not found.

A planned comparison was made between the nondisabled students' performance at standard time ($M = 17.72$) and the LD students' performance at extended time ($M = 12.38$). This was done to examine whether extended time would "level the playing field" and allow the LD students to perform at a level equal to the nondisabled students under standard time conditions. Results indicated that there was still a significant difference between the scores

of the nondisabled group at standard time and the scores of the LD group under extended time conditions that favored the nondisabled group, $t(62) = 4.67$, $p < 0.001$, $d = 1.17$.

Items Attempted. As with items correct, the ANOVA examining attempted NDRT items revealed a significant group by time interaction, $F(1, 62) = 9.69$, $p < .05$, $\eta^2 = .135$, but again, the interaction favored students without disabilities. Group comparisons yielded a significant difference favoring the nondisabled students at standard time, $t(62) = 9.57$, $p < .001$, $d = 2.39$, and this difference became significantly larger at extended time, $t(62) = 12.53$, $p < .001$, $d = 3.13$. Again, the expected differential boost was not found.

Also, the nondisabled students' number of attempted items at standard time ($M = 22.88$) was compared to the LD students' number of attempted items at extended time ($M = 21.25$). A t test found no significant difference, $t(62) = 1.2$, $p > .10$, $d = .06$. This result suggests that the students with LD were able, with 50% extended time, to attempt as many questions as the nondisabled students did at standard time.

Percentage Correct. The percentage of attempted items answered correctly was calculated at both standard and extended time. For this dependent measure, there was no significant interaction of group and time, $F(1, 62) = .001$, $p > .10$, $\eta^2 = .00$, again failing to find differential boost. In addition, there was no main effect of time, but a main effect was found for group, $F(1, 62) = 29.1$, $p < .001$, $\eta^2 = .32$. The nondisabled students demonstrated a higher percentage of correctness, collapsed over time, than the students with LD ($M = 77.1\%$ and 58.78%, respectively).

Discussion

In this study, we examined whether high school students with LDs benefited more from extended time accommodations on a reading comprehension test than their nondisabled classmates. In fact, we found that the nondisabled group benefited significantly more. Also, as expected, students with LDs scored significantly lower on all timed reading comprehension measures (number of attempted, correct, and percentage correct items) than the nondisabled students at both time intervals. Despite the lower performances, students with LDs did show significant improvement in both the number of attempted, as well as correct, items when given extended time. However, even with extended time, their performance did not rise to the level of the nondisabled group's performance at standard time, although it did allow them to attempt the same number of items.

Several limitations of the present study should be considered when interpreting the results. First, although the school district that classified the students with LDs used a standard protocol for doing so, we did not evaluate any of the students ourselves. However, our LD group certainly performed differently from the nondisabled students on most academic measures; yet, they had average nonverbal cognitive ability, which we would expect from students with LDs in the area of reading. Second, the measures in this study were adapted for group administration, and the time limits on the NDRT were modified. Therefore, we did not maintain standardized clinical procedures and could not compare our results to the test norms. Third, our sample was fairly small and ethnically homogeneous, suggesting that

the results may not extend to other populations. Finally, we could not simulate a "high stakes" test environment—that is, students did not have the pressure (or the motivation) that surrounds many standardized assessments. Therefore, the use of a time accommodation may have a different meaning in the context of our study. In a related point, students were informed in advance that they would receive extended time, just like they would if they were being given an actual test accommodation. This could have influenced the way in which students paced themselves and ultimately performed.

Despite these limitations, the results of this study are quite similar to those found by Lewandowski and colleagues (2007) in an ADHD sample. We showed that when ceiling effects are controlled, all examinees benefited from extended time. In fact, nondisabled examinees actually showed greater gains with extended time. At the same time, these results conflict with previous studies that did not control for ceiling effects (compare Sireci et al., 2005). Of course, almost all the extant studies have shown that nondisabled students do benefit, at least somewhat, from extended time, and this is a crucial point. As Zuriff (2000) noted, the rationale for extended testing time for students with disabilities is that they are not performing at their "maximum potential" when they are only allowed a standard amount of time, whereas nondisabled students are performing at their maximum potential under standard conditions. Therefore, when nondisabled students benefit from an accommodation, the appropriateness of that accommodation must be reassessed.

Implications for Practice

Based on Zuriff's (2000) arguments, then, or even under the more liberal differential boost criterion (Fuchs & Fuchs, 2001), one could argue that the results of this study would not support extended time accommodations, because extended time's effects were not specific to students with disabilities. However, from another perspective (namely, advocacy for individuals with disabilities), it is clear that the students with LD had a disadvantage that was unrelated to their cognitive ability: they read more slowly than their nondisabled peers, and they completed fewer test items. In other words, they were significantly impaired relative to their average peers. The extended time did allow these students to access as much of the test as nondisabled peers without outperforming their peers, and that would seem to argue in favor of the accommodation in cases where significant impairment exists.

We would endorse a policy that acknowledges both the lack of differential boost and the importance of accessing test items, namely, a policy of eliminating or diminishing the time element from tests whenever possible. The notion of universal test design (e.g., Thompson, Johnstone, Anderson, & Miller, 2005) suggests that tests should be constructed in a way that is fair to all examinees, making accommodations unnecessary. With the increasing shift to computer-based testing (Parshall, Spray, Kalohn, & Davey, 2002), tests are likely to become more capable of minimizing the effects of time constraints, and classroom teachers' duty to adapt assessments for students with special needs may lead them to consider making tests in which more students can access more of the items. Eventually, then, extended time could become an infrequent accommodation reserved for those rare students with severe impairments and very unusual needs.

Directions for Future Research

Although the present study has implications for practice, more research is necessary to further understand extended time accommodations. First, because these accommodations keep tests from measuring speed, we must learn which tests should be measuring speed and which should not. For instance, it is unlikely that a college-level psychology examination should be measuring speed, that is, the instructor who administers the test is unlikely to be interested in his or her students' speed at answering items; whereas an elementary school teacher who is assessing his or her students' reading skills may wish to consider speed, given the importance of reading fluency in predicting later comprehension skills (e.g., Chard, Pikulski, & McDonagh, 2006). Most admission and credentialing exams include a speed component; yet it is unclear whether speed is conceptually relevant to the skill being tested.

A second topic for future research involves taking a more nuanced approach to investigating differential boost effects—specifically, looking for moderators of these effects. For instance, the present study, as well as that of Lewandowski et al. (2007), suggests that the presence of ceiling effects may determine whether differential boost is found, further suggesting that the speededness of the test (the proportion of the variability in test scores because of variability in number of items attempted; Gulliksen, 1950) is a moderator of differential boost effects. Other possible moderators include the content area and specific academic skills being tested, the response format (e.g., computerized tests vs. paper and pencil), the amount of extended time allotted (e.g., time and one-half vs. double time), and the personality characteristics of students (e.g., levels of test anxiety).

Future research, then, must proceed in at least two directions, one examining the conditions under which the differential boost effects are found and one examining broader questions about which tests should be speeded. Equipped with this knowledge, we will be able to craft test accommodations policy for individuals substantially limited by disability, allowing them to access test items without providing inflated test outcomes in which validity is compromised—only then will we be making accurate evaluations of examinees' skills, leading to the most appropriate instructional, admissions, and credentialing decisions.

References

Barkley, R. A. (1997). *ADHD and the nature of self-control*. New York: Guilford.

Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy. *Remedial and Special Education, 25*, 141-152.

Bracken, B. A., & Naglieri, J. A. (2003). Assessing diverse populations with nonverbal tests of general intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2, pp. 243-274). New York: Guilford.

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny Reading Test, Form H.* Itasca, IL: Riverside.

Chard, D. J., Pikulski, J. J., & McDonagh, S. H. (2006). Fluency: The link between decoding and comprehension for struggling readers. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (pp. 39-61). New York: Guilford.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Court, J. H., & Raven, J. (1995). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York: Guilford.

Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound testing accommodation decisions for students with learning disabilities. *Learning Disabilities Research and Practice*, *16*, 174-181.

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley.

Katsiyannis, A., Zhang, D., Ryan, J. B., & Jones, J. (2007). High-stakes testing and students with disabilities: Challenges and promises. *Journal of Disability Policy Studies*, *18*, 160-167.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.

Lee, D., Reynolds, C. R., & Wilson, V. L. (2003). Standardized test administration: Why bother? *Journal of Forensic Neuropsychology*, *3*(3), 55-81.

Lewandowski, L., Codding, R., Kleinmann, A., & Tucker, K. (2003). Assessment of reading rate in postsecondary students. *Journal of Psychoeducational Assessment, 21*, 134-144.

Lewandowski, L., Lovett, B., Parolin, R., Gordon, M., & Codding, R. (2007). The effects of extended time on mathematics performance of students with and without attention deficit hyperactivity disorder. *Journal of Psychoeducational Assessment, 25*, 17-28.

Mather, N., & Wendling, B. J. (2005). Linking cognitive assessment results to academic interventions for students with learning disabilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (2nd ed., pp. 269-294). New York: Guilford.

Ofiesh, N. S., Hughes, C., & Scott, S. S. (2004). Extended test time and postsecondary students with learning disabilities: A model for decision making. *Learning Disabilities: Research and Practice, 19*, 57-70.

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.

Shaywitz, S. (2003). *Overcoming dyslexia*. New York: Vintage.

Sireci, S. G., Scarpati, S., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*, 457-490.

Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). *Consideration for the development and review of universally designed assessments* (NCEO Technical Report No. 42). Minneapolis, MN: National Center on Educational Outcomes.

Thurlow, M. L., Elliott, J. L., & Ysseldyke, J. E. (2003). *Testing students with disabilities: Practical strategies for complying with district and state requirements* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Thurlow, M. L., Thompson, S. J., & Lazarus, S. S. (2006). Considerations for the administration of tests to special needs students: Accommodations, modifications, and more. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 653-673). Mahwah, NJ: Erlbaum.

Tindal, G., & Fuchs, L. (1999). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: University of Kentucky, Mid-South Regional Resource Center Interdisciplinary Human Development Institute.

Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). *Testing handicapped people*. Boston: Allyn & Bacon.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson tests of achievement* (3rd ed.). Itasca, IL: Riverside.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13*, 99-117.